EURASIP Journal on
Information Security

RESEARCH Open Access

Multi-resolution privacy-enhancing technologies for smart metering

Fabian Knirsch[1,2*], Günther Eibl[1] and Dominik Engel[1]

Abstract

The availability of individual load profiles per household in the smart grid end-user domain combined with non-intrusive load monitoring to infer personal data from these load curves has led to privacy concerns. Privacy-enhancing technologies have been proposed to address these concerns. In this paper, the extension of privacy-enhancing technologies by wavelet-based multi-resolution analysis (MRA) is proposed to enhance the options available on the user side. For three types of privacy methods (secure aggregation, masking and differential privacy), we show that MRA not only enhances privacy, but also adds additional flexibility and control for the end-user. The combination of MRA and PETs is evaluated in terms of privacy, computational demands, and real-world feasibility for each of the three method types.

Keywords: Smart meter, Homomorphic encryption, Masking, Differential privacy, Multiple resolutions, Smart grid

1 Introduction

Intelligent energy systems and so-called smart grids, change the way electricity is generated, distributed, and used. The widespread roll-out of smart meters is one of the consequences. Such smart meters record energy consumption in a specified granularity (usually the time between readings is between 1 and 15 min, cf. Table 10 in [1]) and have the ability to transmit these load curves in a specified interval (e.g., once a day). Therefore, this involves a considerable amount of information that needs to be processed and analyzed. Smart grids further demand accurate and fine-grained data on network status, as well as a detailed analysis of load profiles from customers [2]. This is crucial for applications such as billing with dynamic pricing, demand response, and network monitoring.

However, it has been shown that personal information on the end-user can be inferred from fine-grained load curves [3, 4], and this has led to privacy concerns (e.g., [5]). This also implies some severe privacy threats such as the identification of customer presence at home, customer habits, and even the customer position when using electric vehicles [6]. In [7, 8], the authors show the impact of resolutions on privacy and that information can be deduced even at comparably low frequencies.

The accuracy of the inferred information is directly connected to the available resolution of the load data. A number of methods have been proposed to balance the need for privacy with the information needed for correct operation of smart grids. Two types of approaches show high potential to resolve this issue: (i) privacy-aware aggregation of encrypted load curves; and (ii) representation of load curves in multiple resolutions, each associated with different access levels.

1.1 Privacy-aware aggregation

Approaches for privacy-aware aggregation can again be divided into three categories: protocols using masking [9, 10], protocols using secure aggregation by homomorphic encryption [11, 12], and protocols using differential privacy [13, 14]. In this paper, the focus is put on the application of multi-resolution load curve representation in combination with secure aggregation protocols.

Privacy-enabling encryption for smart meter data by the use of homomorphic encryption is suggested by, [11, 12, 15, 16], allowing the aggregation of encrypted signals, also termed “secure signal processing”. A recent

*Correspondence: fabian.knirsch@en-trust.at
[1] Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Salzburg University of Applied Sciences, Urstein Süd 1, 5412 Puch bei Hallein, Austria
[2] Department of Computer Sciences, University of Salzburg, Jakob-Haringer-Str. 2, 5020 Salzburg, Austria

overview of secure signal processing, covering four proposals for privacy-preserving smart metering aggregation is given in [17]. Protocols that are using masking for aggregating data have been proposed by [9, 10, 18]. Masking approaches aim to hide individual contributions by additive noise, but still produce a valid aggregate. Differential privacy follows a similar approach, where contributions are hidden in a noisy aggregate that fulfills some statistical properties. Differential privacy is adapted for applications in the smart grid by, e.g., [13, 19–21].

1.2 Multiple resolutions

Approaches of this type suggest to represent load curve data in multiple resolutions, where each resolution can be used for a different purpose — e.g., low resolution for billing — and is therefore disclosed to selected parties only, e.g., [22]. Using the wavelet transform in order to produce an integrated bitstream supporting multiple resolutions has been proposed by [23]. Combined with conditional access, i.e., different encryption keys for each resolution [24], this wavelet-based representation allows user-centric privacy management: access can be granted or revoked for each resolution. Access to high resolutions, which are privacy-sensitive, may be reserved to a small number of trusted entities only, whereas resolutions of medium granularity may be provided more freely, e.g., to contribute to network stability (in exchange for lower energy prices or other incentives). An approach combining multiple resolutions and direct user control for smart metering is shown in [25]. The combination of MRA with homomorphic encryption, which is also one of the topics in this paper, has been discussed in [26].

1.3 Contribution

In this paper, a set of three privacy-preserving smart metering data aggregation methods that combine the two types of approaches, namely, multi-resolution representation and (i) homomorphic encryption; (ii) masking; and (iii) differential privacy, is proposed. This improves the capabilities for managing privacy requirements, as the combination of "traditional" privacy-enhancing methods with multi-resolution representation significantly increases the choices available for both system operator and end-user. We further contribute the sketch of a protocol for distributing keys and for providing distinct resolutions to different parties. Access control does not relate to the aggregated signal as a whole anymore, but access can be granted on the aggregate on each resolution *individually*. This is an important feature, as it allows to grant access to participants in the smart grid system, based on their roles and the functions they have to fulfill. Each role can be assigned access to the aggregate on the minimum resolution necessary to fulfill the functions associated with this role.

The combination of MRA with homomorphic encryption has previously been proposed in [26]. This paper extends the previous work by applying multi-resolution techniques to masking and differential privacy. A comprehensive presentation, discussion, and evaluation of multi-resolution representation in combination with widely used PETs is given.

The rest of this paper is structured as follows: in section Multi-resolution PETs the application scenario and common definitions are introduced. In section Background, background is presented on wavelets for the multi-resolution representation of load curves as well as on the three privacy-enhancing technologies (PETs) homomorphic encryption, masking, and differential privacy. Sections Multi-resolution secure aggregation, Multi-resolution masking, and Multi-resolution differential privacy describe each of these PETs individually and propose the combination of these approaches with wavelets. In section Evaluation, the security features of the proposed protocols, as well as cost and complexity are discussed, and further, the system is evaluated with respect to real-world applicability on the basis of a prototypical implementation. Section Conclusions summarizes this paper and gives an outlook to future work.

2 Multi-resolution PETs

While homomorphic encryption, simple masking, and differential privacy are efficient methods for the *spatial* reduction of resolution, *temporal* aggregation is not sufficiently covered with any of these approaches. Temporal resolution of time series can be reduced by subsequently applying a number of filters. When —for instance — applying an appropriate low-pass filter to a time series, all frequencies above the cutoff frequency are omitted, which results in a signal with less information. This is effectively performed by applying the wavelet transform, in particular the Haar wavelet, to a series of values.

2.1 Application scenario

Smart meter data has a wide range of applications, such as in-house monitoring, billing, network monitoring, and demand response. As pointed out in [2, 8], data resolution depends on the use case and has an impact on the privacy, i.e., the information the recipient can gain from that data. In the following, we introduce three typical application scenarios and motivate the need for multi-resolution PETs and their flexibility with respect to spatial and temporal resolution.

1. Settlement and profiling. In the energy market electricity generators and electricity suppliers trade at a wholesale marketplace. The arrangement of payments among these parties is called settlement [2, 9]. Profiling is used for determining forecasts and

training models, e.g., in the UK this is based on half-hourly meter data from a representative sample of households [2]. Both applications thus require data in a comparably low resolution, but spatially aggregated over a number of households.

2. Network monitoring. Network monitoring is used for detecting outages and peaks and thus maintaining the stability of the power grid. A detailed monitoring of power consumption, voltage levels and phase shifts is an important feature for network operators. For monitoring purposes, data at a high temporal resolution but with little spatial resolution is required.
3. Billing. Billing requires meter data in a low temporal resolution (e.g., one value per month or year), however, on a per household or on a per meter basis, hence not spatially aggregated at all. In future applications, dynamic pricing might also require more fine-grained data [27]. Multi-resolution PETs enable the provision of load profiles in certain resolutions depending on the particular use case.

2.2 Topology

For data aggregation in the smart grid, a number of different topologies are proposed, such as star topologies (e.g., [20, 28]), ring topologies (e.g., [10]), and tree topologies (e.g., [16]). In any case, the smart meters generate a time series of values that is either sent to a dedicated collector node or to a data concentrator, which is responsible for aggregating these measurements. Or the smart meters aggregate in a hop-by-hop manner, i.e., a smart meter sends its measurement to its successor or parent node where this measurement is combined with its own value. The data concentrator and the last smart meter, respectively, forward the aggregated measurements to one or more recipients (in the following referred to as *aggregators*). Each aggregator receives data in a different spatio-temporal resolution depending on the role of the recipient and the needed granularity. Figure 1 shows examples for star and tree topologies. For the protocols presented in this paper, the aggregation method (direct or hop-by-hop) is not restricted and either of these approaches can be used.

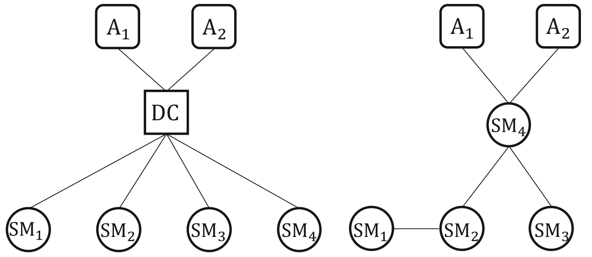

Fig. 1 Examples of two different topologies (*star*, *left* and *tree*, *right*). In the star topology, the data concentrator (DC) collects measurement values from the smart meters (*SM*) and forwards the aggregated values to the aggregators (*A*). In the tree topology measuring values are aggregated in a hop-by-hop manner by the smart meters

2.3 Problem statement and definitions

Given a number of smart meters SM$_i$, for $i = 1 \ldots N$, one or more aggregators A_k, for $k = 1 \ldots M$, and a trusted third party (TTP), each meter i measures a time series of values, i.e., at time t it measures $m_{i,t}$. In this paper, a series of values measured by a meter i is denoted as m_i. In order to protect customer privacy, the sum of the energy consumption for all smart meters should be provided to the aggregator. The following restrictions and requirements apply (aggregator oblivious): (i) no aggregator can gain any information about individual contributions; (ii) each aggregator can only unmask a valid sum up to the time resolution $r \leq R$ (with R as the maximum resolution) that is intended to be revealed for this aggregator. Hence, the aggregator is considered to be untrusted. In practice, the smart meters can be considered to be physically arranged in either a tree or a ring topology. Logic topologies may defer and may depend on the concrete protocol. For homomorphic encryption and masking, the TTP is needed to provide the keys (pk^r, sk^r) and the key shares (key^r), respectively, to the smart meters and aggregators.

For this paper, we assume that there is a sufficient underlying secure communication infrastructure, i.e., the bidirectional and reliable exchange of information and the secure distribution of keys is given as well as authenticated communication among participants is guaranteed by, e.g., AES [29] and X.509 certificates [30]. We further assume all devices to be tamper-proof, i.e., the meter value itself cannot be manipulated.

3 Background

In this section, we briefly review the existing work on multi-resolution representation, homomorphic encryption, masking, and differential privacy.

3.1 Wavelet-based representation

A wavelet transform starts with the original load curve $m = (m_1, m_2, \ldots, m_T)$, which denotes a series of values. Each step splits the original load curve into a high-pass component h and a low-pass components l. If the wavelet transform is performed recursively in d steps, this is denoted as $W_d(m)$. In each step q, for $q = 1, \ldots d$, half of the data (the highpass data) h_q are stored as the wavelet coefficients (subband) of scale q, and the next step is performed for the low-pass data. At the end of the transformation, the final subband h_d consists of a fraction of 2^{-d} samples compared to the original load curve. The higher the scale q, the lower the time resolution $r := d - q$. Reindexing, and introducing the notation $h^r = h_{d-q}$, at the end of the transformation one obtains a sequence $h = (l_0, h_1, \dots, h_d)$.

The synthesis step of the inverse wavelet transform W^{-1} starts with the lowest resolution $r = 0$. To get the next higher resolution of the signal, the next higher resolution subband is needed, so that in a series of d steps one finally obtains the original load curve (since we only consider lossless transformations). In order to provide a signal m^r with maximum resolution r, only r synthesis steps must be performed and only the subbands with resolution $r \leq R$, i.e., $m^r = (l_0, h_1, \dots, h_r)$, are needed. Denoting the selection of the r highest resolutions as a function T_r, this can be written as

$$m^r = W^{-1}\left(T_r(W(m))\right). \tag{1}$$

This selection can be realized in practice by replacing the high-pass subbands with zeros, i.e., applying $T_r(\cdot)$ to a sequence $W(m) = (l_0, h_1, \dots, h_r, \dots, h_{d-1}, h_d)$ yields a sequence $T_r(W(m)) = (l_0, h_1, \dots, h_r, 0, \dots, 0)$. This limits, after applying the inverse wavelet transform, the resolution of the signal. Making the signal available at the needed resolution instead of the full resolution increases privacy because less (personal) information can be deduced [8].

In [23], a variety of wavelet filters regarding their utility for the multi-resolution representation of load curves was evaluated. Only lossless transformations are useful in the context of smart metering. The Haar wavelet filter preserves the average over all resolutions, which is an important property for many use cases. Using the lifting implementation of the Haar wavelet, the transformation can be realized efficiently.

The lifting steps for the forward transform with the Haar wavelet have been formulated by [31]. As the original Haar wavelet uses real coeffcients, it is ill-suited for use with homomorphic encryption. Therefore, for the combination of PETs, a modified version of the Haar wavelet is used that only produces integer values for the transformed load curve. While this is generally not an issue for masking and differential privacy, we still use the modified version for all PETs. A detailed description of the Haar wavelet lifting scheme can be found in [23]. Note that the average of the original series is still preserved over all resolutions for the modified Haar filter:

$$\forall r: \sum_{t=0}^{T} m_t = 2^{-r} \sum_{t=0}^{T} m_t^r. \tag{2}$$

3.2 Additive homomorphic encryption

Following previous proposals [11, 12, 15] for this work, the Paillier cryptosystem [32] is employed. This additive homomorphic cryptosystem has the following important property, which is called the *additive property*:

$$D\left(E(m_1)E(m_2) \bmod n^2\right) = (m_1 + m_2) \bmod n. \tag{3}$$

This property means that the decryption of the product of the *ciphertexts* is the sum of the original plaintext messages.

In a practical setting, the network is assumed to have tree-like connections. Each smart meter sends its measured load in encrypted form to its parent node. The parent smart meter multiplies the obtained encrypted signals with its own encrypted signal and in turn sends this product to its parent node. Finally, the aggregator multiplies the obtained signals and decrypts the product. Due to the additive homomorphic property, the result is the sum of the measurements. With E and D denoting Pailler encryption and decryption, this can be stated as

$$D\left(\prod_i E(m_i) \bmod n^2\right) = \sum_i m_i \mod n. \tag{4}$$

Privacy is preserved because of the distributed way of processing. Smart meters only have the plaintext information of their own messages, because they cannot decrypt the messages they get. The aggregator can decrypt messages, but, as it receives the product of the individual ciphertexts, it can only decrypt the sum of the load curves.

3.3 Masking

Masking refers to the obfuscation of individual contributions, such that the summation of load profiles over a number of households yields the correct sum, but no individual contribution is traceable. This is achieved by adding for each SM_i at time t a random share s_i in the range $1, \dots, \kappa - 1$ to the meter value m_i. This results in a masked meter value $\tilde{m}_i = m_i + s_i \mod \kappa$. The set of random shares is constructed in such a way that

$$\sum_i \tilde{m}_i = \sum_i (m_i + s_i) = \sum_i m_i \text{ (all mod } \kappa), \tag{5}$$

hence, the shares cancel each other out upon summation.

Principally, smart meters calculate the masked value $\tilde{m}_i$ and submit this value to an aggregator. Once the aggregator has received all masked values, it can calculate the unmasked sum. If a single value is missing, the secret shares will not cancel each other out, and neither the aggregate nor any individual contribution can be reconstructed.

Kursawe et al. [9] present a number of methods for constructing such shares that meet the requirement for untraceability of individual contributions: (i) aggregation protocols for determining the sum as described above; and (ii) comparison protocols that require the aggregator already knows an (at least) approximate sum. For our purpose, we focus on the low-overhead protocol from the first group which has already been used in practical implementations [33]. For the low-overhead protocol, all smart meters hold a public key $\mathrm{pk}_i = g^{X_i}$ with X_i as a secret key and $g \in \mathbb{G}$ as a generator of a group satisfying the computational Diffie-Hellman assumption [34]. Each SM_i is given the set of all public keys and computes a set of $N-1$ shared keys by $K_{i,j} = H(\mathrm{pk}_j^{X_i})$ with $j = 1 \ldots N$.

As described in [9], for each meter value at time t each SM_i creates a random share by

$$s_i = \sum_{k \neq i} (-1)^{b(i,j)} H(K_{i,j} || t), \tag{6}$$

where $b(i,j)$ returns 1 if $j < i$ and 0 otherwise, and $H : \{0,1\}^* \rightarrow \mathbb{G}$ is a hash function mapping its input to an element of $\mathbb{G}$. This term in Eq. 6 results in $+H(K_{i,j}||t)$ for $b(j,i) = 0$ and $-H(K_{i,j}||t)$ for $b(j,i) = 1$. Summing up this values assures that all s_i cancel each other out pairwise since $K_{i,j} = K_{j,i}$ because of $g^{X_i X_j} = g^{X_j X_i}$. This is shown for N smart meters at one point in time t in Table 1, where the rows represent k, and the columns represent i for values from 1 to N. Summing up the resulting terms in each row yields the random share $s_{i,t}$.

3.4 Differential privacy

Differential privacy is a privacy definition that defines privacy of a function f by an indistinguishability property of the function result. In this paper, the function is the time series of the sum of different smart meter measurements $f(t) = \sum_{i=1}^{N} T_r\left(W(m_i)\right)$. However, note that here the noise is added to the selected resolutions (operator T_r) in the wavelet domain and not in the original domain. The aim is that by examining a perturbed result $\tilde{f}(t)$, one cannot distinguish whether a single person's entry is contained or not. Since the noise is only added to the needed resolutions $\leq r$, only a small amount of noise is added. More formally, two neighboring datasets $\mathcal{D}$ and $\mathcal{D}'$ that differ in the entries of a single person/household only are considered. The function mechanism $\tilde{f}$ is then ϵ-differentially private, if for a small privacy parameter $\epsilon > 0$

$$\Pr[\tilde{f}(\mathcal{D}) = y] \leq \exp(\epsilon) \Pr[\tilde{f}(\mathcal{D}') = y] \, . \tag{7}$$

Table 1 In the method proposed by [9], shares cancel each other out pairwise, since $s_i + s_{N-i} = 0$

j/i	1	2	3	...	N
1		$-H(K_{2,1}\|\|t)$	$-H(K_{3,1}\|\|t)$	...	$-H(K_{N,1}\|\|t)$
2	$+H(K_{1,2}\|\|t)$		$-H(K_{3,2}\|\|t)$	...	$-H(K_{N,2}\|\|t)$
3	$+H(K_{1,3}\|\|t)$	$+H(K_{2,3}\|\|t)$		...	$-H(K_{N,3}\|\|t)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	$+H(K_{1,N}\|\|t)$	$+H(K_{2,N}\|\|t)$	$+H(K_{3,N}\|\|t)$	...	
$\sum_j$	$s_{1,t}$	$s_{2,t}$	$s_{3,t}$	...	$s_{N,t}$

The columns correspond to N smart meters, the last row is the share created for each smart meter for one point in time t

While differential privacy is a theoretically appealing definition with nice properties (e.g., a function is differentially private under postprocessing), it is achieved by perturbing the function result with Laplacian noise $\tilde{f}(t) = f(t) + n_t$ [19], where each noise value n_t is independently and identically sampled from a Laplacian distribution $n_t \sim \mathrm{Lap}_\lambda$ (the parameter λ must be set using the sensitivity of the function f [19]). As a drawback, the function result is not exact and can be useless if the number of entries in the dataset is too small.

More specifically, according to the Theorem of Dwork [19], the L_p sensitivity of a function $f : D^n \rightarrow \mathbb{R}^d$ is the smallest number $S_p(f)$ such that for two neighboring datasets x and x'

$$S_p(f) = \underset{x,x'}{\operatorname{argmax}} \, \|f(x) - f(x')\|_p. \tag{8}$$

The most common mechanism that achieves differential privacy is the Laplace mechanism $\mathcal{M}_L$ that perturbs the output of f by adding noise from a Laplace distribution having the density

$$\mathrm{Lap}_\lambda(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right), \tag{9}$$

in a non-interactive way, yielding

$$\mathcal{M}_L(x, f(\cdot), \epsilon) = f(x) + (Y_1, \ldots, Y_k), \quad \text{with } Y_l \overset{i.i.d.}{\sim} \mathrm{Lap}_\lambda. \tag{10}$$

An important theorem states that the Laplace mechanism is ϵ-differentially private if the parameter λ is chosen by

$$\lambda = \frac{S_1(f)}{\epsilon}. \tag{11}$$

The resulting noise does not need to be added directly to the function result. It can also be added in a distributed manner [20, 35] when each contributing party i adds i.i.d. noise $G_{\lambda,N}$ defined by

$$\Pr[G_{\lambda,N} = x] = \mathrm{G}^1_{1/N,\lambda}(x) - \mathrm{G}^2_{1/N,\lambda}(x), \tag{12}$$

where G^1 and G^2 are two i.i.d. gamma distributions with identical shape parameter $1/N$ and scale parameter λ. Then

$$\Pr[n_t = x] = \sum_{i=1}^{N} G_{\lambda,N}(x) = \mathrm{Lap}_\lambda(x). \tag{13}$$

4 Multi-resolution secure aggregation

In this section, the combination of the wavelet transform with homomorphic encryption is presented. The principal scheme is shown in Fig. 2. First, the basic approach with only one aggregator is presented, and second, it is shown

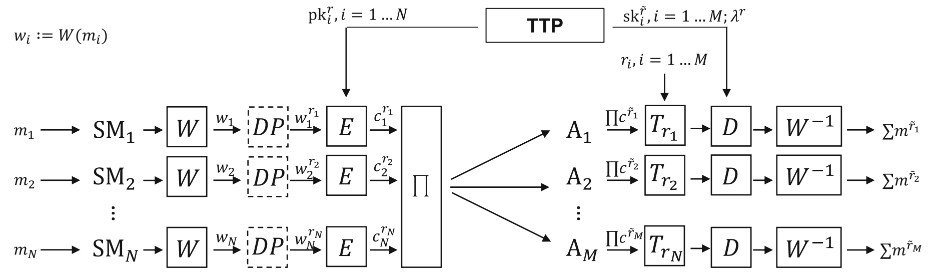

Fig. 2 MRA aggregation scheme for secure aggregation with homomorphic encryption and multiple aggregators. The component DP representing the addition of Gamma distributed noise is only needed if additionally differential privacy wants to be achieved. In this figure $w_i := W(m_i)$ denotes the wavelet transform of m_i

that this approach can easily be extended to multiple aggregators.

4.1 Principal secure aggregation scheme

Homomorphic encryption is applied to each resolution separately with a different pair of keys $(\mathrm{pk}_r, \mathrm{sk}_r)$ for each resolution r. The resulting signal m is the sum of all signals m_i (each of which has a maximum resolution of R) at resolution $r \leq R$, whereby, $W(\cdot)$ denotes a wavelet transformation. The collector node can perform aggregation (i.e., multiply) in the encrypted domain, i.e., it does not have any keys. This ensures that the aggregating node cannot get information about the loads of its children, e.g., by divisions.

4.2 Basic approach

The basic approach covers a number of smart meters and a single aggregator. Writing the principal scheme mathematically yields the following calculation of the ciphertext c

$$c = \prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) \mod n^2. \tag{14}$$

The ciphertext c is decrypted by the aggregator by

$$m = W^{-1}\left(D\left(c\right) \mod n\right). \tag{15}$$

Using this procedure, the wavelet transformation is compatible with homomorphic encryption, i.e., the property that the message m equals the sum of the messages is preserved (choosing $r = R$). Even more, choosing $r \leq R$, the decrypted message m equals the sum of the messages of resolution r:

$$m = W^{-1}\left(D\left(\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) \mod n\right)\right) = \sum_i m_i^r \mod n. \tag{16}$$

The aggregator gets the product of the encrypted messages and can therefore not extract any information about the individual messages. However, it can calculate the sum of the messages which is the information needed, e.g., for load forecasting. Note again that the product of the ciphertexts is calculated in either a distributed way by the smart meters or by a data concentrator and not by the aggregator (see section Topology). The number n must be chosen depending on the desired security level. It further determines the aggregation group size, since $\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) < n^2$ and $D\left(\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right)\right) < n$. In section Space considerations, the issue of aggregation group sizes is discussed in detail. For the sake of readability the modulus parts of the calculations are omitted in the following proof.

Proof Without loss of generality, two messages are considered. To simplify the analysis the notation $y_i := T_r\left(W\left(m_i\right)\right)$ is used, so $E\left(T_r\left(W\left(m_i\right)\right)\right) = E(y_i)$. The aggregator calculates the signal $W^{-1}(D(c))$. Using the fact that the ciphertext c is the product of the individual ciphertexts and the homomorphic encryption property leads to

$$\begin{aligned} W^{-1}\left(D\left(c\right)\right) &= W^{-1}\left(D\left(c_1 c_2\right)\right) \\ &= W^{-1}\left(D\left(E\left(y_1\right)E\left(y_1\right)\right)\right) \\ &= W^{-1}(y_1 + y_2) \end{aligned} \tag{17}$$

Substituting the y_i using the linearity of the wavelet transform and the definition of m^r yields

$$\begin{aligned} W^{-1}\left(D\left(c\right)\right) &= W^{-1}\left(T_r\left(W\left(m_1\right)\right) + T_r\left(W\left(m_2\right)\right)\right) \\ &= W^{-1}\left(T_r\left(W\left(m_1\right)\right)\right) + W^{-1}\left(T_r\left(W\left(m_2\right)\right)\right) \\ &= m_1^r + m_2^r \end{aligned} \tag{18}$$

So in general for N different messages and ciphertext $c = \prod_i c_i$, the desired property (16)

$$W^{-1}\left(D\left(c\right)\right) = \sum_{i=1}^{N} m_i^r. \tag{19}$$

is obtained. □

4.3 Multiple aggregators

An example use-case scenario is the use of aggregated load information for energy monitoring by the network operator as, e.g., suggested by [17]. The approach proposed here adds an additional layer of flexibility by making the aggregates available at different resolutions and only grant access to parties on the resolutions they need to fulfill a specific task. In combination with suitable key management, this approach implements the "need-to-know" principle of access for aggregated signals. The secure aggregation scheme presented above can be extended to support multiple aggregators. Each aggregator receives data in a certain resolution. This is easily achieved by encrypting with different keys at the collector node.

5 Multi-resolution masking

In this section, the multi-resolution masking approach is presented. The principal scheme is shown in Fig. 3. After briefly recapitulating the principal masking scheme, first, the basic approach for one aggregator is presented and second, this approach is extended to multiple aggregators receiving data in different resolutions. The latter is especially useful for application scenarios such as settlement and profiling, where different parties should be provided information in different resolutions.

5.1 Principal masking scheme

Each smart meter SM_i calculates at each time $t = 0 \ldots T$ a masked value $\tilde{m}_{i,t}$ by adding a random share $s_{i,t}$ to its measured value $m_{i,t}$. Upon spatial aggregation, the shares s_i cancel each other out and the aggregator receives an unmasked sum. Note that in the following, operations involving masking of type $a + b \mod \kappa$ are written as $a + b$, i.e., the modulo parts are omitted for the sake of brevity and readability.

This approach can be enhanced by allowing to reduce the temporal resolution of the signal. Even more, a number of different resolutions can be provided within the same bitstream, and the key for a certain resolution is only given to the aggregator. This is achieved by applying a wavelet transform to the signal. Hence, even if the aggregator is given the full load curve data, it can only unmask the bitstream up to the resolution for which it holds the key share.

5.2 Basic approach

The basic approach describes spatio-temporal masking with one aggregator.

5.2.1 Initialization

TTP agrees with all smart meters in the group $G = \{\mathrm{SM}_1, \ldots, \mathrm{SM}_N\}$ on providing a resolution r of a total of T values to an aggregator A.

5.2.2 Masking

Simultaneously, all SM_i and TTP calculate a random share $s_{i,t}$ for $t = 0 \ldots T$, as described for the principal masking above. Each smart meter now holds a set of shares s_i and TTP holds a key share key.

All SM_i now calculate a series of masked values $\tilde{m}_i = W(m_i) + s_i$ and submit this series to A. TTP calculates the key share key^r for the resolution r of its key share by $T_r(\mathrm{key})$ and submits this to A. Note that the wavelet transform is only applied to the metered value and before adding the random share.

5.2.3 Aggregation

After receiving both, the shares from all smart meters and the key share A can calculate the aggregated sum over all smart meters at a time resolution r by

$$\sum_i m_i^r = W^{-1}\left(T_r\left(\sum_i \tilde{m}_i\right) + \mathrm{key}^r\right). \tag{20}$$

If the aggregator attempts to retrieve any resolution $r^+ > r$, the result will be noisy and useless. However, the aggregator may reconstruct arbitrary resolutions $r^- \leq r$ from the data.

Proof Proof that reconstructing a resolution r^+ for a key with resolution r will be noisy. The aggregator receives an aggregation of the masked meter values

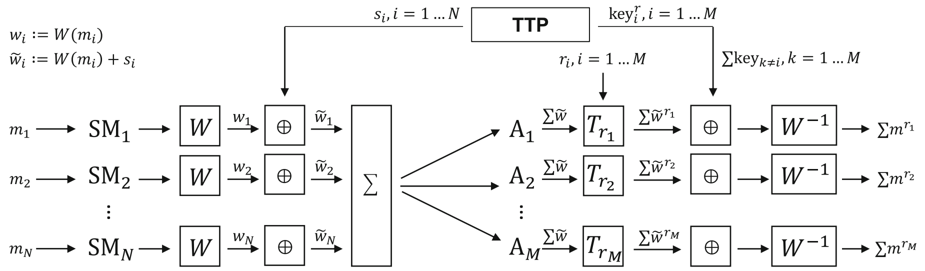

Fig. 3 MRA aggregation scheme for masking and multiple aggregators. In this figure, $w_i := W(m_i)$ denotes the wavelet transform of m_i, and $\tilde{w}_i := W(m_i) + s_i$ denotes the masked values of w_i

$$\sum_i \tilde{m}_i = \sum_i \left(W\left(m_i\right) + s_i \right), \tag{21}$$

and a key share $\text{key}^r = T_r(\text{key})$ for some resolution r. Applying this function $T_r(\cdot)$ to a series of values replaces the high-pass components by zeros. The key share and the random shares for masking have the property that

$$\sum_i s_i + \text{key} = 0, \tag{22}$$

but that the key share for a particular resolution r yields

$$\sum_i s_i + \text{key}^r \neq 0, \tag{23}$$

since the high-pass components are set to zero in the key share and do not cancel out the corresponding components in the sum of the shares. Therefore,

$$W^{-1}\left(\sum_i \left(W\left(m_i\right) + s_i \right) + \text{key} \right) = \sum_i m_i, \tag{24}$$

and

$$W^{-1}\left(\sum_i \left(W\left(m_i\right) + s_i \right) + \text{key}^r \right) \neq \sum_i m_i. \tag{25}$$

However, after applying the function $T_r(\cdot)$ with the same parameter r to the equation, this yields Eq. 20 which is the correct result for this particular resolution r. Note that the wavelet transform is recursively applied to the resulting low-pass band, i.e., any resolution $r^- < r$ can be retrieved, since applying $T_r(\cdot)$ to the key share only replaces the high-pass components by zero, and only the low-pass components remain for reconstructing the signal.

□

Note that this scheme fulfills both of our initial requirements: (i) individual contributions are masked, and the aggregator cannot gain any information without having all the values from all $\text{SM}_i \in G$; and (ii) the highest resolution that is accessible for the aggregator is determined by the resolution of the key share.

5.3 Multiple aggregators

The scheme we present in the following extends the basic approach with multiple aggregators that receive data in different resolutions. Extending the scheme requires more overhead and communication than for the secure aggregation. A simple approach would be to have multiple bitstreams in multiple resolutions for each aggregator. The advantage of the MRA approach is, however, to have all the information for different resolutions in a single bitstream where no data expansion occurs. Therefore, a different key share for every recipient is created with the tradeoff of distributing an aggregate of $M - 1$ key shares in addition to the actual key share.

5.3.1 Initialization

For the enhanced scheme supporting multiple aggregators $L = \{A_1, \ldots, A_M\}$, a TTP agrees with all smart meters in the group $G = \{\text{SM}_1, \ldots, \text{SM}_N\}$ on providing a resolution r_k of a total of T values to each aggregator $A_k \in L$.

5.3.2 Masking

As in the basic scheme, all smart meters SM_i calculate a random share $s_{i,t}$ for $t = 0 \ldots T$. Again, each smart meter now holds a set of shares s_i, calculates the series of masked values $\tilde{m}_i = W(m_i) + s_i$. and submits this series to all aggregators $A_k \in L$. TTP calculates a total of M (number of aggregators) key shares $\text{key}_1, \ldots, \text{key}_M$. For each key share $k = 1 \ldots M$, TTP further calculates the resolution r_k by $\text{key}_k^{r_k} = T_{r_k}(\text{key}_k)$ and submits this to A_k. It further submits the sum of all other key shares $\sum_{i \neq k} \text{key}_i$ to A_k.

5.3.3 Aggregation

After receiving both, the shares from all smart meters and the set of key shares, each $A_k \in L$ can calculate the aggregated sum over all smart meters at a time resolution r_k by

$$\sum_i m_i^{r_k} = W^{-1}\left(T_r\left(\sum_i \tilde{m}_i \right) + \text{key}_k^{r_k} + T_r\left(\sum_{i \neq k} \text{key}_i \right) \right). \tag{26}$$

As for the basic approach, both of our initial requirements are fulfilled: (i) individual contributions are masked, and none of the aggregators can gain any information without having all the values from all $\text{SM}_i \in G$ and the sum of all other key shares $\sum_{i \neq k} \text{key}_i$; and (ii) the highest resolution that is accessible for each aggregator is determined by the resolution of the individual key share. These requirements are fulfilled due to the properties of the masking approach as introduced in section Masking and formally shown in section Basic approach.

5.4 Proof of correctness

In the following, it is shown that applying the wavelet transform to a meter value and masking can be combined in order to provide a certain resolution only. This proof is — for simplicity and without loss of generality — for a single smart meter and a single aggregator. The proof also applies to multiple smart meters and multiple aggregators. The only difference is that instead of a single meter value, share and key, respectively, a (spatially) aggregated sum of values is used. For multiple aggregators, the sum of all other key shares is also required as shown in the previous section.

Proof Proof for a single aggregator that it is receiving m_i^r at the end of the above masking scheme. Starting from

$$\underbrace{W\left(m_i\right)+s_i}_{\mathrm{SM}_i}+\underbrace{\text{key}}_{TTP}=\underbrace{W\left(\hat{m}_i\right)}_{A}, \tag{27}$$

where the braces indicate what the smart meter and the TTP calculate and what the aggregator receives at the end of the protocol, $T_r \circ W^{-1}$ is applied on both sides of the equation:

$$W^{-1}\left(T_r\left(W\left(m_i\right)+s_i+\text{key}\right)\right)=W^{-1}\left(T_r\left(W\left(\hat{m}_i\right)\right)\right). \tag{28}$$

Due to the linearity of both, the wavelet transform and the function $T_r(\cdot)$ this is equivalent to

$$W^{-1}\left(T_r\left(W\left(m_i\right)\right)\right)+W^{-1}\left(T_r\left(s_i\right)+T_r\left(\text{key}\right)\right)=\hat{m}_i^r. \tag{29}$$

Substituting $m_i^r = W^{-1}\left(T_r\left(W\left(m_i\right)\right)\right)$, $s_i^r = T_r\left(s_i\right)$ and $\text{key}_i^r = T_r\left(\text{key}\right)$ results in

$$m_i^r + W^{-1}\left(s_i^r + \text{key}_t^r\right) = \hat{m}_i^r. \tag{30}$$

Given the property of masking, shares cancel each other out by $s_i^r + \text{key}^r = 0$, and therefore $m_i^r + W^{-1}\left(0\right) = \hat{m}_i^r$.

This is obviously equivalent to $m_i^r = \hat{m}_i^r$, i.e., the aggregator only receives a certain resolution m_i^r of the original meter value m_i. □

6 Multi-resolution differential privacy

In this section, it is shown that the wavelet approach can be combined with an additional differential privacy method. The benefit of this approach is an additional ϵ-differential privacy guarantee (Eq. 7) for the resulting *aggregated* signal.

6.1 Combining wavelets and differential privacy

Combining differential privacy in a *distributed* way with wavelets, only guarantees differential privacy for the sum, but not for the individual signals. Therefore, combining differential privacy with wavelets alone, would not enhance privacy so that the combination with homomorphic secure aggregation is needed. Similar to the masking approach, the combination with the differential privacy method requires a distributed addition and later summation of random values. The scheme is described in Fig. 2 and leads, using the additive homomorphic property of the encryption, to the following intermediate result.

$$\begin{aligned}\tilde{f} &= W^{-1}\left(D\left(\prod_{i=1}^{N} c_i^{r_i}\right)\right) = W^{-1}\left(\sum_{i=1}^{N} w_i^{r_i}\right)\\ &= W^{-1}\left(\sum_{i=1}^{N}\left(T_{r_i}\left(W\left(m_i\right)\right)+G_{\lambda,N}\right)\right).\end{aligned} \tag{31}$$

The additional use of homomorphic encryption is not the only difference to masking. In contrast to masking, the random values are drawn *independently* from each other from a non-uniform probability distribution $G_{\lambda,N}$, denoted as block *DP* in Fig. 2. Due to Eq. 13, these distributedly generated probability distributions sum up to the Laplacian distribution which is needed for differential privacy of the *aggregate* profile

$$\tilde{f} = W^{-1}\left(\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right)+\text{Lap}_\lambda\right). \tag{32}$$

Thus, if the noise parameter λ is chosen such that $\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right)$ is ϵ-differentially private, due to the postprocessing property of differential privacy, also $\tilde{f} = W^{-1}\left(\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right)+\text{Lap}_\lambda\right)$ is ϵ-differentially private. Finally, using the linearity of W,

$$\tilde{f} = \sum_{i=1}^{N} m_i^{r_i} + W^{-1}(\text{Lap}_\lambda), \tag{33}$$

is shown to be a perturbed function of the smoothed consumption sum. This smoothed consumption sum is ϵ-differentially private, if the Laplacian noise is set in the right manner. Therefore, in principle, the wavelet decomposition is compatible with differential privacy.

Another difference to the presented masking scheme is that the noise is added to the *restricted* wavelet values instead of the unrestricted values $W(m_i)$ (Eq. 32). However, since several different resolutions occur, setting the right amount of noise λ is not trivial and remains a task for future research. First preliminary steps in that direction show that it is possible to derive a choice for λ which, however, only provides differential privacy for a single resolution r. With such a noise differential privacy can only be provided for a single resolution r and, due to the post-processing property, all coarser solutions.

6.2 Choice of parameter λ

In this subsection, we show how the parameter λ must be chosen by proving the following theorem.

Theorem (choice of λ): the presented algorithm is ϵ-differentially private, if (i) W is a tight frame; and (ii) parameter λ is chosen as

$$\lambda = \frac{\sqrt{\mathcal{R}}}{\epsilon} \operatorname*{argmax}_{m_{i,\cdot}} \left\|m_{i,\cdot}\right\|_2, \tag{34}$$

where $\mathcal{R}$ denotes the number of coefficients up to resolution $r = d - q$.

Note that $\mathcal{R}$ consists of a fraction of 2^{-q} samples compared to the original load curve. The smaller the resolution r, the smaller the λ, and therefore the added noise is chosen.

Proof First, the situation of this algorithm must be properly mapped into the differential privacy setting. Note that the term $\tilde{w}^r$ of the algorithm can be rewritten as

$$\tilde{w}^r = \sum_{i=1}^{N} \left(T_r\left(W(m_i)\right) + \mathrm{G}^1_{1/n,\lambda_\epsilon}(x) - \mathrm{G}^2_{1/n,\lambda_\epsilon} \right). \quad (35)$$

Due to the divisibility property, the sum of the Gamma-distributions yield the Laplace distribution. Thus, we have

$$\begin{aligned} \tilde{m}^r = W^{-1}(\tilde{w}^r) &= W^{-1}\left(\mathcal{M}_L(m,f(\cdot),\epsilon)\right) \\ &= W^{-1}\left(f(m) + \mathrm{Lap}_{\lambda_\epsilon}\right). \end{aligned} \quad (36)$$

If we manage to prove differential privacy for our choice of *f*, the proof is finished since a function applied to a differentially private mechanism can not destroy the differential privacy property (closure under post-processing property of differential privacy). Therefore, if $\tilde{w}^r$ is ϵ-differentially private this also holds for $\tilde{m}^r = W^{-1}(\tilde{w}^r)$.

In order to prove differential privacy for our choice of *f*, we will show that the choice of λ ensures that it is at least as big as the one of theorem Differential privacy, whose application then proves differential privacy. Since m and m' differ in a single household's entry, we can write without loss of generality that $m = (m_{1,\cdot}, \ldots, m_{N,\cdot}, m_{N+1,\cdot}) = (m', m_{N+1,\cdot})$. Since in theorem Differential privacy, a 1-norm is needed instead of a 2-norm, first the transition is done using the inequality

$$\|x\|_2 \geq \|x\|_1/\sqrt{\mathcal{R}}.$$

Note that this inequality can itself be proven by applying the Cauchy-Schwarz inequality to $\langle \mathbb{1}, |x| \rangle$. Together with the linearity of T_r and W, this yields

$$\begin{aligned} \|f(m) - f(m')\|_1 &\leq \sqrt{\mathcal{R}}\|f(m) - f(m')\|_2 \\ &= \sqrt{\mathcal{R}} \left\| \sum_{i=1}^{N+1} T_r\big(W(m_{i,\cdot})\big) - \sum_{i=1}^{N} T_r\big(W(m_{i,\cdot})\big) \right\|_2 \\ &= \sqrt{\mathcal{R}} \left\| T_r\left(W\left(\sum_{i=1}^{N+1} m_{i,\cdot} - \sum_{i=1}^{N} m_{i,\cdot} \right)\right) \right\|_2 \\ &= \sqrt{\mathcal{R}} \| T_r\left(W\left(m_{N+1,\cdot}\right)\right) \|_2 \end{aligned} \quad (37)$$

The restriction to a smaller resolution is equivalent to setting the higher resolutions to zero. Therefore, the restriction T_r decreases the norm while the wavelet transformation does not change it due to our restriction of using only transformations with the tightness property

$$\begin{aligned} \|f(m) - f(m')\|_1 &\leq \sqrt{\mathcal{R}} \| W\left(m_{N+1,\cdot}\right) \|_2 \quad (38) \\ &= \sqrt{\mathcal{R}} \|m_{N+1,\cdot}\|_2. \quad (39) \end{aligned}$$

Finally, this equation directly yields

$$\begin{aligned} \lambda &= \frac{\sqrt{\mathcal{R}}}{\epsilon} \operatorname*{argmax}_{m_{N+1,\cdot}} \|m_{N+1,\cdot}\|_2 \quad (40) \\ &\geq \frac{1}{\epsilon} \operatorname*{argmax}_{m,m'} \|f(m) - f(m')\|_1 \quad (41) \\ &= \frac{S_1(f)}{\epsilon}. \quad (42) \end{aligned}$$

Thus, theorem Differential privacy can be applied and proves differential privacy for *f*. □

7 Evaluation

In this section, we evaluate the proposed PETs in combination with MRA with respect to the security features, cost and complexity, and real-world applicability.

7.1 Applications

In this paper, multi-resolution secure aggregation has been introduced for both single aggregator and multiple aggregators. Given the building blocks of the additive homomorphic Paillier cryptosystem, masking, and differential privacy in combination with the wavelet transform, a scheme can be constructed that allows to encrypt different resolutions with different keys while maintaining a single bitstream. In section Application scenario, three typical application scenarios for smart grid have been introduced: (i) settlement and profiling; (ii) network monitoring; and (iii) billing.

Settlement and profiling require data in a comparably low resolution, but spatially aggregated over a number of households for determining forecasts and training models. Network monitoring, by contrast, still works with the aggregate, but requires a much higher temporal resolution. Both homomorphic encryption and masking can be used for aggregating over a number of smart meters, e.g., from households connected to the same substation or participants belonging to the same consumption group (residential/industrial). By adding the ability to selectively decrypt a subset of multiple resolutions, the same aggregated bitstream, but with different keys, can be provided to both the utility provider for forecasts and model training and the network operator for network monitoring. This reduces the overhead for managing and transferring various bitstreams simultaneously to distinct recipients. While network monitoring might require very high accuracy (e.g., voltage levels must remain in a narrow band), for settlement and profiling, customer privacy can be even enhanced by adding differential privacy in order to prevent the detection of the presence of a single household in the aggregate, while at the same time providing a certain guaranteed ϵ-differential privacy level. While differential privacy is a compelling approach due to this property, it is not suitable for applications that require the exact aggregate.

Billing and dynamic pricing will require data at high resolutions and generally not aggregated. Further, differential privacy is not a desired property for billing. However, if in a dynamic pricing scenario, data in different granularity is needed over the day (e.g., a stable night tariff and more dynamic tariffs at noon), the multi-resolution approach allows to dynamically adjust the level of granularity of the provided meter data.

7.2 Security analysis

In this section, a security analysis of the proposed PETs is conducted. We consider an honest-but-curious adversarial model, meaning the adversary follows the protocols but tries to gain additional information.

Secure signal processing: for MRA with secure signal processing, an honest-but-curious aggregator will not learn any information. Due to the additive homomorphic property of the cryptosystem, even at collector nodes, all operations are performed in the encrypted domain, and the aggregator can only decrypt the sum.

Masking with single aggregator: for a total number of smart meters $N > 1$ and a single aggregator $M = 1$, the masking scheme preserves full privacy in terms of spatial resolution, and it preserves full privacy with respect to temporal resolution. Given exactly one aggregator $M = 1$, the *basic approach* for multi-resolution masking is applied. A receives a set of N masked values and a single key share. By combining both, A can calculate the sum at a particular resolution. For spatial aggregation, privacy is preserved by the scheme proposed by Kursawe et al. [9], i.e., the individual measurements are masked, and the random shares cancel each other out upon summation. The temporal resolution is limited by the resolution of the key share. The privacy preserving feature of this approach has been discussed in detail in section Basic approach. Section Proof of correctness includes the proof of correctness for the masking approach in combination with wavelets.

Masking with multiple aggregators: for a total number of smart meters $N > 1$ and a total number of aggregators $M = 2$, the *multiple aggregators approach* for multi-resolution masking is applied. Each aggregator $A_k, k = \{1, 2\}$ receives a set of N masked values, an individual key share key_1^r and key_2^r, respectively and the sum of the keys of all other aggregators $\sum_{i \neq 1} \text{key}_i^r = \text{key}_2^r$ and $\sum_{i \neq 2} \text{key}_i^r = \text{key}_1^r$, respectively. Therefore, each aggregator additionally holds the other aggregators share in full resolution and thus privacy in terms of temporal aggregation is not given anymore.

For a setting with $M > 2$ privacy is preserved, as the key shares of all other aggregators are hidden in the sum. Therefore, the above limitation for $M = 2$ does not apply, since $\sum_{i \neq k} \text{key}_i^r \neq \text{key}_i^r$ for any $i, k \in \{1 \ldots M\}$. This means that holding all keys except for one does not yield a valid key. This assures that the aggregator cannot learn anything beyond the resolution of the key, which is formally shown in section Multiple aggregators.

Differential privacy: it is a proven property of differential privacy, that the aggregator has no means to decrease privacy of the *aggregated* signal by any kind of postprocessing. If differential privacy would be combined with wavelets only, the aggregator could, however, inspect a *single* smart meter's consumption profile. A single profile is only protected by Gamma-distributed noise which does not provide differential privacy. Therefore, the mechanism achieving differential privacy must include a way to protect the summation operation by using a secure aggregation scheme, e.g., as described in section Multi-resolution secure aggregation.

7.3 Space considerations

When using homomorphic encryption for aggregation, the modulus n determines the amount of data that can be stored within one encrypted packet. Let us denote the number of bits needed to represent a wavelet coefficient by $\bar{m}$, and the number of values used for the wavelet transform by T. The sum of two coefficients will take up $\bar{m} + 1$ bits of space. More generally, the sum of u coefficients requires $\lceil \log_2(u) + \bar{m} \rceil$ bits. If encrypting each wavelet coefficient individually, i.e., using T encryptions, the modulus of n bits allows to sum up a total of $u \leq 2^{n-\bar{m}}$ wavelet coefficients, since $n = \lceil \log_2(u) + \bar{m} \rceil$, i.e., u represents the total number of wavelet coefficients from household measurement values that can be aggregated for a given modulus.

Setting $T = 256$, $n = 1024$ and $\bar{m} = 16$, this allows for the aggregation of more than $2 \cdot 10^{303}$ households but requires 256 encryptions. However, in practice such large aggregation groups are not needed. Instead of encrypting each coefficient individually, the available space of n bits can be exploited better when using data packing [36]. Values are shifted to a certain bit range, such that a number of values can be packed within a single encryption. The available space is therefore split into p packets of fixed size n', i.e., $p = \frac{n}{n'}$. This allows for $n' = \lceil \log_2(u') + \bar{m} \rceil$ a number of $u' \leq 2^{n'-\bar{m}}$ wavelet coefficients per packet, and a total of $u' \cdot p$ wavelet coefficients of household measurement values per encryption. This results in only $T' = \frac{T}{p}$ encryptions.

Setting $n = 1024$, $\bar{m} = 16$ and $n' = 32$, this results in $p = 32$ packets and still allows to aggregate up to 65536 households, but with only a fraction ($T' = 8$) of the number of encryption operations compared to the above approach where each coefficient is encrypted separately. In practice, these values have to be chosen with respect to the number of households that will be aggregated.

7.4 Cost and complexity

Both methods secure signal processing with the Paillier cryptosystem[1] and masking have been implemented together with the wavelet transform. The proof of concept implementation is built on Oracle Java 1.8 and is tested on a HP Z230 workstation with 8 GB RAM and an Intel Xeon CPU (3.4 GHz). Results are shown in Table 2: each value represents the execution time for a single load curve consisting of 96 values for the wavelet transform combined with different encryption settings and masking, averaged over 400 load curves with 100 encryptions/additions for masking each (acquisition of the load curve and key generation as well as precalculating the masking shares are not considered in the timing results). WAV denotes the wavelet transform only, without any encryption or masking applied. AES denotes the wavelet transform followed by encryption with the symmetric AES cipher with a 256 bit key for each subband. HYB denotes hybrid encryption, which adds RSA 2048 bit public key encryption of the AES keys with a different public key for each subband. PAI-n denotes Pailler encryption with a module of n bits and a different key for each subband. For practical applications and according to [37], a module of at least 2048 bits should be chosen. Finally, MA denotes the masking of values.

It can be seen that by using a lifting implementation, the computational overhead of the wavelet transformation is negligible compared to the encryption step. Homomorphic encryption comes at the cost of a significant increase in computational overhead compared to that of conventional encryption. The results show that the computational demands grow exponentially with the module size. Although the used implementation of Paillier is not optimized and could be improved considerably in terms of efficiency, it is clear that running homomorphic encryption on smart meter hardware will provide a challenge: while AES encryption only takes 1.25 ms, for the used (non-optimized) implementation, Paillier encryption with a 2048 bit module of a load curve with 96 values takes approximately 52 s. Further, it can be seen that masking is highly efficient in terms of computation time when compared to encryption, however, at the cost of losing the entire aggregate when a single smart meter fails.

Table 2 Execution time t in milliseconds and standard deviation σ for transforming/encrypting/masking a single load curve (average over 400 load curves with 100 encryptions each)

	WAV	AES	HYB	PAI-2048	PAI-4096	MA
t	< 0.001	0.07	0.7	5,219	38,700	< 0.001
σ	< 0.001	0.02	0.01	25.4	51	< 0.001

8 Conclusions

The approaches proposed in this paper allow to get both temporal and spatial aggregation by combining the wavelet transform with homomorphic encryption, masking, and differential privacy. In this paper, it has been shown that it is possible to combine homomorphic encryption, masking, and differential privacy with the Haar wavelet transform. Furthermore, a protocol has been sketched for addressing different aggregators with different resolutions of the measured time series while still maintaining a certain level of privacy. For masking, future work will focus on a scheme that is more error-resilient and still yields the correct result even if a subset of smart meters fail.

Endnote

[1] Building on the implementation by Kun Liu http://www.csee.umbc.edu/~kunliu1/research/Paillier.html

Acknowledgements
The financial support by the Austrian Federal Ministry of Science, Research and Economy and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Federal State of Salzburg is gratefully acknowledged.

Authors' contributions

This paper was written by FK (40%), GE (30%), and DE (30%). The detailed contributions are as follows: the Abstract was written by FK (100%). The Introduction was written by FK (80%) and DE (20%). The Background section was written by FK (40%), DE (40%), and GE (20%). The Multi-resolution Secure Aggregation section was written by DE (100%). The Multi-resolution Masking section was written by FK (100%). The Multi-resolution Differential Privacy section was written by GE (100%). The Evaluation section was written by FK (60%), DE (30%), and GE (10%). Conclusions and Outlook were written by FK (100%). The figures were created by FK (100%). Measurements for homomorphic encryption were performed by DE (100%). All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 18 August 2016 Accepted: 3 March 2017
Published online: 20 March 2017

References

1. European Commission (2014). Cost-benefit analyses and state of play of smart metering deployment in the EU-27. Technical report. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014SC0189&from=EN.
2. McKenna, E, Richardson, I, Thomson, M (2012). Smart meter data: balancing consumer privacy concerns with legitimate applications. *Energy Policy*, *41*, 807–14.
3. Hart, GW (1992). Nonintrusive appliance load monitoring. *Proc. IEEE*, *80*(12), 1870–91.
4. Molina-Markham, A, Shenoy, P, Fu, K, Cecchet, E, Irwin, D (2010). Private memoirs of a smart meter, In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building. BuildSys '10* (pp. 61–6). New York: ACM.
5. Lisovich, M, Mulligan, D, Wicker, S (2010). Inferring personal information from demand-response systems. *IEEE Secur. Priv*, *8*(1), 11–20.

6. Knirsch, F, Engel, D, Frincu, M, Prasanna, V (2015). Model based assessment for balancing privacy requirements and operational capabilities in the smart grid, Innovative smart grid technologies conference (ISGT), 2015 IEEE Power & Energy Society, In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT2015)* (pp. 1–5). Washington: IEEE.
7. Eibl, G, & Engel, D (2014). Influence of data granularity on nonintrusive appliance load monitoring, In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)* (pp. 147–51). Salzburg: ACM.
8. Eibl, G, & Engel, D (2015). Influence of data granularity on smart meter privacy. *IEEE Trans. Smart Grid*, *6*(2), 930–9.
9. Kursawe, K, Danezis, G, Kohlweiss, M (2011). Privacy-friendly aggregation for the smart grid, In *Privacy Enhanced Technology Symposium* (pp. 175–91). Berlin Heidelberg: Springer.
10. Gomez Marmol, F, Sorge, C, Petrlic, R, Ugus, O, Westhoff, D, Martinez Perez, G (2013). Privacy-enhanced architecture for smart metering. *Int. J. Inf. Secur*, *12*(2), 67–82.
11. Li, F, Luo, B, Liu, P (2010). Secure Information Aggregation for Smart Grids Using Homomorphic Encryption, In *Proceedings of First IEEE International Conference on Smart Grid Communications* (pp. 327–332). Gaithersburg: IEEE.
12. Erkin, Z, & Tsudik, G (2012). Private computation of spatial and temporal power consumption with smart meters, In *Proceedings of the 10th International Conference on Applied Cryptography and Network Security. ACNS'12* (pp. 561–77). Berlin: Springer.
13. Rastogi, V, & Suman, N (2010). Differentially private aggregation of distributed time-series with transformation and encryption, In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*. Indianapolis: ACM, Conference.
14. Danezis, G, Kohlweiss, M, Rial, A (2011). *Differentially private billing with rebates* Vol. 6958 LNCS, (pp. 148–62). Berlin: Springer.
15. Garcia, F, & Jacobs, B (2011). Privacy-friendly energy-metering via homomorphic encryption. In J Cuellar, J Lopez, G Barthe, A Pretschner (Eds.), *Security and Trust Management. Lecture Notes in Computer Science*, 6710 (pp. 226–38). Berlin: Springer.
16. Li, F, & Luo, B (2012). Preserving data integrity for smart grid data aggregation, In *Third International Conference on Smart Grid Communications (SmartGridComm) 2012* (pp. 366–71). Tainan: IEEE.
17. Erkin, Z, Troncoso-pastoriza, JR, Lagendijk, RL, Perez-Gonzalez, F (2013). Privacy-preserving data aggregation in smart metering systems: an overview. *IEEE Signal Proc. Mag*, *30*(2), 75–86.
18. Biselli, A, Franz, E, Coutinho, MP (2013). Protection of consumer data in the smart grid compliant with the German smart metering guideline, In *Proceedings of the First ACM Workshop on Smart Energy Grid Security. SEGS '13* (pp. 41–52). New York: ACM.
19. Dwork, C, McSherry, F, Nissim, K, Smith, A (2006). Calibrating noise to sensitivity in private data analysis, In *Theory of Cryptography* (pp. 265–84). Berlin: Springer.
20. Acs, G, & Castelluccia, C (2011). I have a DREAM! (DiffeRentially privatE smArt Metering), In *Proc. Information Hiding Conference* (pp. 118–132). Berlin Heidelberg: Springer.
21. Shi, E, Chow, R, Chan, THH, Song, D, Rieffel, E (2011). Privacy-preserving aggregation of time-series data, In *Proc. NDSS Symposium 2011*. San Diego: Internet Society.
22. Efthymiou, C, & Kalogridis, G (2010). Smart grid privacy via anonymization of smart metering data, In *Proceedings of First IEEE International Conference on Smart Grid Communications* (pp. 238–43). Gaithersburg: IEEE.
23. Engel, D (2013). Wavelet-based load profile representation for smart meter privacy, In *Proc. IEEE PES Innovative Smart Grid Technologies (ISGT'13)* (pp. 1–6). Washington: IEEE.
24. Peer, CD, Engel, D, Wicker, SB (2014). Hierarchical key management for multi-resolution load data representation, In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (pp. 926–32). Venice: IEEE.
25. Engel, D, & Eibl, G (2016). Wavelet-based multiresolution smart meter privacy. *IEEE Trans. Smart Grid*, *PP*(99), 1–12.
26. Engel, D, & Eibl, G (2013). Multi-resolution load curve representation with privacy-preserving aggregation, In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013* (pp. 1–5). Copenhagen: IEEE.
27. Jawurek, M, Johns, M, Kerschbaum, F (2011). Plug-in privacy for smart metering billing, In *Privacy Enhancing Technologies (PETS)* (pp. 192–210). Berlin Heidelberg: Springer.
28. Erkin, Z (2015). Private data aggregation with groups for smart grids in a dynamic setting using CRT, In *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. Rome: IEEE.
29. (2001). National Institute of Standards and Technology (NIST), Specification for the advanced encryption standard (AES).
30. ITU-T (2012). Recommendation ITU-T X.509 – Information technology – open systems interconnection – the directory: public-key and attribute certificate frameworks.
31. Daubechies, I, & Sweldens, W (1998). Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl*, *4*(3), 247–69.
32. Paillier, P (1999). Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In J Stern (Ed.), *Advances in Cryptology — EUROCRYPT '99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, May 2–6, 1999 Proceedings. Lecture Notes in Computer Science*, 1592 (pp. 223–38). Berlin: Springer.
33. Defend, B, & Kursawe, K (2013). Implementation of privacy-friendly aggregation for the smart grid, In *Proceedings of the First ACM Workshop on Smart Energy Grid Security - SEGS '13* (pp. 65–74). Berlin: ACM, Conference.
34. Diffie, W, & Hellman, M (1976). New directions in cryptography. *IEEE Trans. Inf. Theory*, *22*(6), 644–654.
35. Kotz, S, Kozubowski, TJ, Krzysztof, P (2001). *The Laplace distribution and generalizations*. Basel: Birkhäuser Basel.
36. Erkin, Z, Veugen, T, Toft, T, Lagendijk, RL (2012). Generating private recommendations efficiently using homomorphic encryption and data packing. *IEEE Trans. Inf. Forensic. Secur*, *7*(3), 1053–66.
37. Barker, E, Barker, W, Burr, W, Polk, W, Smid, M (2012). *Division, Computer Security NIST Special Publication 800-57, Recommendation for Key Management (Revision 3)*: NIST.